



HAL
open science

La Géométrie pour l'Unicité, la Parcimonie et l'Appariement des Estimateurs Pénalisés

Patrick Tardivel, Ulrike Schneider

► **To cite this version:**

Patrick Tardivel, Ulrike Schneider. La Géométrie pour l'Unicité, la Parcimonie et l'Appariement des Estimateurs Pénalisés. 53èmes Journées de Statistique de la SFDS, Jun 2022, Lyon, France. hal-03945880

HAL Id: hal-03945880

<https://hal.science/hal-03945880>

Submitted on 18 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LA GÉOMÉTRIE POUR L'UNICITÉ, LA PARCIMONIE ET L'APPARIEMENT DES ESTIMATEURS PÉNALISÉS

Patrick Tardivel ¹ & Ulrike Schneider ²

¹ *Institut de Mathématiques de Bourgogne UMR 5584, CNRS Université de Bourgogne
F-21000 Dijon, France, patrick.tardivel@u-bourgogne.fr*

² *TU Wien, ulrike.schneider@tuwien.ac.at*

Résumé. Durant la présentation nous donnerons une condition nécessaire et suffisante pour l'unicité d'un estimateur des moindres carrés pénalisés dont le terme de pénalité est une norme polyédrique. Nos résultats couvrent de nombreuses méthodes incluant les estimateurs OSCAR, SLOPE et LASSO ainsi que la méthode de poursuite de base qui est une extension du LASSO. Cette condition d'unicité est géométrique et implique l'espace vectoriel engendré par les lignes de la matrice de planification ainsi que les faces de la boule unité de la norme duale coupées par cet espace. Des résultats théoriques sur la parcimonie des estimateurs LASSO et poursuite de base sont déduits de cette condition via la caractérisation des vecteurs signes accessibles pour ces deux méthodes.

Mots-clés. Unicité, LASSO, poursuite de base, vecteur signe accessible.

Abstract. During the talk we will give a necessary and sufficient condition for the uniqueness of a penalized least squares estimator whose penalty term is a polyhedral norm. Our results cover many methods including the OSCAR, SLOPE and LASSO estimators as well as the related method of basis pursuit. The geometrical condition for uniqueness involves how the row span of the design matrix intersects the faces of the dual norm unit ball. Theoretical results on sparsity by LASSO and basis pursuit estimators are deduced from this condition via the characterization of accessible sign vectors for these two methods.

Keywords. Uniqueness, LASSO, basis pursuit, accessible sign vector.

1 Condition nécessaire et suffisante pour l'unicité

De nombreux estimateurs des moindres carrés pénalisés tels que le LASSO (Least Absolute Shrinkage and Selection Operator), clustered LASSO, OSCAR (Octogonal Shrinkage and Clustering Algorithm for Regression) ou SLOPE (Sorted L One Penalized Estimation)... sont solutions du problème (1) donné ci-dessous. Soit $X \in \mathbb{R}^{n \times p}$, $y \in \mathbb{R}^n$ et $\|\cdot\|$ une norme sur \mathbb{R}^p dont la boule unité est un polytope. L'ensemble $S_{X,\|\cdot\|}(y)$ des solutions du problème des moindres carrés pénalisés est défini par

$$S_{X,\|\cdot\|}(y) := \operatorname{Argmin}_{b \in \mathbb{R}^p} \frac{1}{2} \|y - Xb\|_2^2 + \|b\|. \quad (1)$$

On peut remarquer que $S_{X, \|\cdot\|}(y)$ n'est jamais vide mais $S_{X, \|\cdot\|}(y)$ n'est pas toujours un singleton. En statistique y représente la réponse aléatoire d'un modèle de régression linéaire ainsi il est pertinent de donner une condition sur la matrice X garantissant l'unicité du problème (1) pour tout $y \in \mathbb{R}^n$.

Le Théorème 1 fournit une condition sur X pour que l'ensemble $S_{X, \|\cdot\|}(y)$ soit un singleton pour tout élément $y \in \mathbb{R}^n$. Rappelons que pour une norme $\|\cdot\|$ sur \mathbb{R}^p , la norme duale $\|\cdot\|^*$ est définie par $\|x\|^* = \sup_{s \in \mathbb{R}^p: \|s\| \leq 1} s^t x$.

Théorème 1 *Soit $X \in \mathbb{R}^{n \times p}$ et $\|\cdot\|$ une norme pour laquelle la boule unité est un polytope. Il existe $y \in \mathbb{R}^n$ pour lequel l'ensemble $S_{X, \|\cdot\|}(y)$ n'est pas un singleton si et seulement si l'espace vectoriel $\text{im}(X^t)$ coupe une face de la boule unité de la norme duale B^* dont la dimension est strictement inférieure à $\dim(\ker(X))$.*

Le Théorème 1 prouvé dans l'article Schneider et Tardivel (2022) généralise le Theorème 15 de Ewald et Schneider (2020) traitant de l'unicité du LASSO. Le Théorème 1 s'étend également au cas où le terme de pénalité est une gauge polyédrique à valeurs finies (Tardivel et al. (2021)). Les Figures 1 et 2 illustrent le Théorème 1.

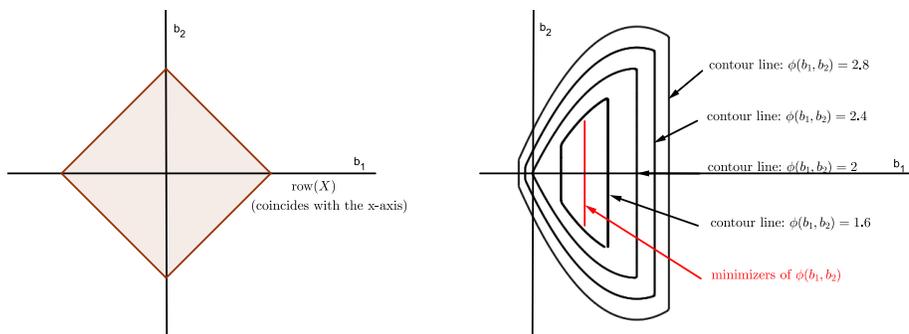


Figure 1: Lorsque $X = (1 \ 0)$, les courbes de niveau de la fonction $\phi(b_1, b_2) = 0,5(2 - b_1)^2 + \max\{|b_1|, |b_2|\}$ illustrent que l'ensemble $S_{X, \|\cdot\|_\infty}(2)$ n'est pas un singleton.

Notons que cette condition d'unicité est en pratique très faible. En effet, en posant μ la mesure de Lebesgue sur $\mathbb{R}^{n \times p}$ et $\|\cdot\|$ une norme sur \mathbb{R}^p dont la boule unité est un polytope alors, l'ensemble suivant est négligeable :

$$\mu(\{X \in \mathbb{R}^{n \times p} : \exists y \in \mathbb{R}^n \text{ tel que } S_{X, \|\cdot\|}(y) \text{ n'est pas un singleton}\}) = 0.$$

Néanmoins, dans certains cas le problème LASSO n'a pas une unique solution. Par exemple, lorsque $X \in [-1, 1]^{n \times p}$ est une matrice où $p > n$ et ayant une ligne avec au

moins $n + 1$ composantes dans $\{-1, 1\}$ alors, d'après le Théorème 1, il existe $y \in \mathbb{R}^p$ pour lequel $S_{X,\lambda\|\cdot\|_1}(y)$ n'est pas un singleton. Une telle matrice est fournie dans le jeu de données "gisette". En particuliers, Dupuis et Vaïter (2019) ont montré que pour la matrice X et le vecteur y de ce jeu de données $S_{X,\lambda\|\cdot\|_1}(y)$ n'est pas un singleton dès que $\lambda > 0$ est suffisamment petit.

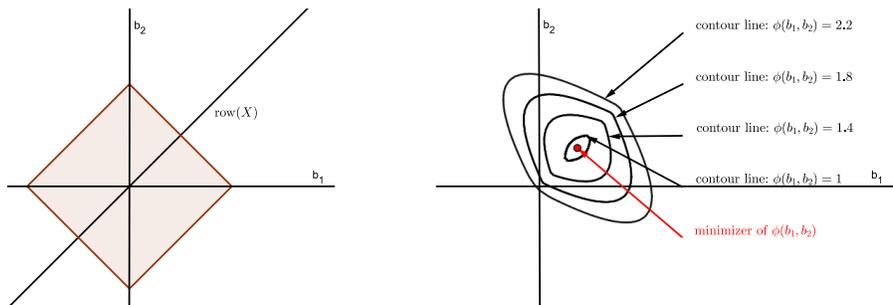


Figure 2: Lorsque $X = \begin{pmatrix} 1 & 1 \end{pmatrix}$, les courbes de niveau de la fonction $\phi(b_1, b_2) = 0,5(2 - b_1 - b_2)^2 + \max\{|b_1|, |b_2|\}$ illustrent que l'ensemble $S_{X,\|\cdot\|_\infty}(2)$ est un singleton.

Le problème poursuite de base introduit ci-dessous peut être vu comme une extension du problème d'optimisation LASSO où le terme de pénalité est $\lambda\|\cdot\|_1$ avec $\lambda > 0$ infiniment petit. L'ensemble $S_{X,\text{pb}}(y)$ des solutions du problème poursuite de base est défini par

$$S_{X,\text{pb}}(y) = \text{Argmin } \|b\|_1 \text{ sous la containte } Xb = y.$$

Une condition nécessaire et suffisante pour l'unicité du problème poursuite de base est donnée au Théorème 2.

Théorème 2 *Soit $X \in \mathbb{R}^{n \times p}$. Il existe $y \in \text{im}(X)$ pour lequel $S_{X,\text{pb}}(y)$ n'est pas un singleton si et seulement si $\text{im}(X^t)$ coupe une face du cube unité $[-1, 1]^p$ dont la dimension est strictement inférieure à $\dim(\ker(X))$.*

Notons qu'une matrice en position générale (condition introduite par Dossal (2012) et Tibshirani (2013) garantissant l'unicité des problèmes LASSO et poursuite de base) est une matrice satisfaisant la condition donnée au Théorème 2. Les Figures 3 et 4 illustrent le Théorème 2.

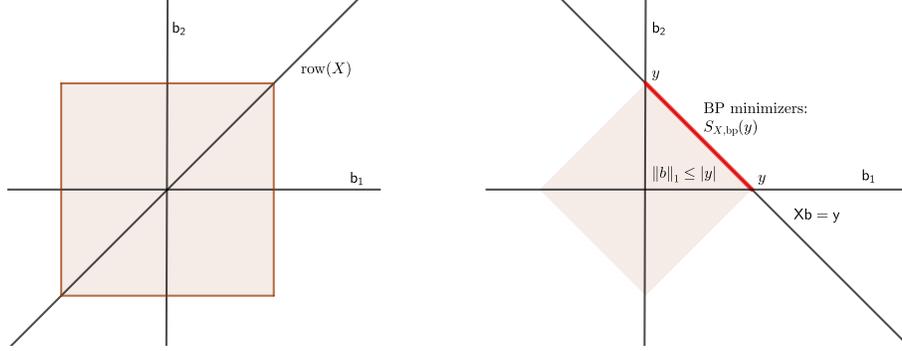


Figure 3: Lorsque $X = (1 \ 1)$, la figure montre que l'ensemble $S_{X,\text{pb}}(y)$ n'est pas un singleton pour un certain $y \in \mathbb{R}$.

2 Vecteur signe accessible pour le LASSO et pour la méthode poursuite de base

La notion de vecteur signe accessible a été introduite pour la première fois par Sepehri and Harris (2017) pour l'estimateur LASSO sous l'hypothèse d'unicité. La définition suivante généralise légèrement la notion de vecteur signe accessible donnée dans Sepehri and Harris.

Définition 1 Soit $X \in \mathbb{R}^{n \times p}$, $\sigma \in \{-1, 0, 1\}^p$ et $\lambda > 0$. On dit que σ est un vecteur signe accessible pour LASSO (ou poursuite de base), s'il existe $y \in \mathbb{R}^n$ et $\hat{\beta} \in S_{X,\lambda\|\cdot\|_1}(y)$ (ou s'il existe $y \in \text{im}(X)$ et $\hat{\beta} \in S_{X,\text{pb}}(y)$), tels que $\text{signe}(\hat{\beta}) = \sigma$.

On pose $F(\sigma)$ la face du cube unité définie par

$$F(\sigma) = E_1 \times \cdots \times E_p \text{ avec } E_j = \begin{cases} \{\sigma_j\} & \text{si } |\sigma_j| = 1 \\ [-1, 1] & \text{si } \sigma_j = 0 \end{cases}.$$

Théorème 3 Soit $X \in \mathbb{R}^{n \times p}$, $\lambda > 0$ et $\sigma \in \{-1, 0, 1\}^p$.

Caractérisation géométrique : Le vecteur de signe σ est accessible pour LASSO (ou poursuite de base) si et seulement si $\text{im}(X^t)$ coupe la face $F(\sigma)$.

Caractérisation analytique : Le vecteur de signe σ est accessible pour LASSO (ou poursuite de base) si et seulement si l'implication suivante est vraie : $Xb = X\sigma \Rightarrow \|b\|_1 \geq \|\sigma\|_1$.

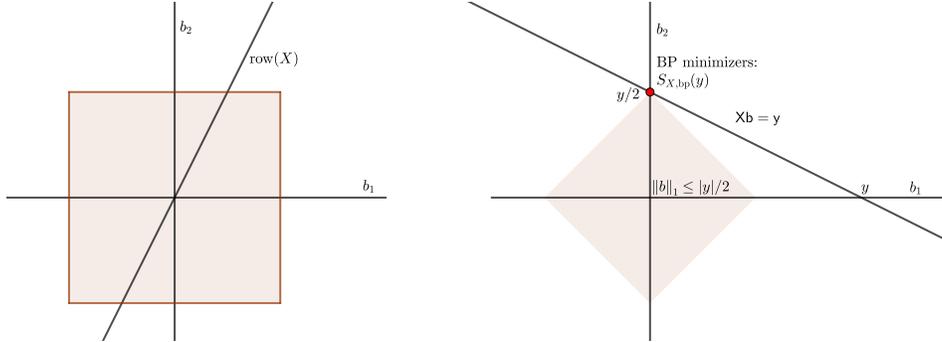


Figure 4: Lorsque $X = (1 \ 2)$, la figure montre que l'ensemble $S_{X,\text{lp}}(y)$ est un singleton.

Dans le cadre du modèle de régression linéaire, la condition d'accessibilité est nécessaire et suffisante pour recouvrir les signes des coefficients de régression (positif, nul ou négatif) par le LASSO seuillé, la méthode poursuite de base seuillée (Tardivel et Bogdan (2022)) et par la méthode poursuite de justice seuillée (Descloux et al. (2020)). Un résultat similaire au Théorème 3, donné dans l'article de Schneider et Tardivel (2022), illustre les propriétés d'appariement et de parcimonie des estimateurs SLOPE et OSCAR.

Remerciements

l'Institut de Mathématiques de Bourgogne (IMB) bénéficie du soutien de l'EIPHI Graduate School (contrat ANR-17-EURE-0002).

Bibliographie

- Descloux, P. and Boyer, C. and Josse, J. and Sportisse, A. and Sardy, S (2020). Robust Lasso-Zero for sparse corruption and model selection with missing covariates, *arXiv*.
- Dossal, C. (2012). A necessary and sufficient condition for exact sparse recovery by ℓ_1 minimization, *Comptes Rendus Mathématique*, 350, 1-2, 117–120.
- Dupuis, X. Vaiter, S. (2019). The geometry of sparse analysis regularization, *arXiv*.
- Ewald, K. and Schneider, U. (2020). On the distribution, model selection properties and uniqueness of the Lasso estimator in low and high dimensions, *Electronic Journal of Statistics*, 14, 1, 944–969.
- Schneider, U. and Tardivel, P. (2022). The Geometry of Uniqueness, Sparsity and Clustering in Penalized Estimation, *Journal of Machine Learning Research*, 23, 331, 1–36.

Sepehri, A. and Harris, N. (2017). *The accessible lasso models*, *Statistics*, 51, 4, 711–721.

Tardivel, P.J.C. and Bogdan, M. (2022). On the sign recovery by least absolute shrinkage and selection operator, thresholded least absolute shrinkage and selection operator, and thresholded basis pursuit denoising, *Scandinavian Journal of Statistics*, 49, 4 1636–1668.

Tardivel, P.J.C and Skalski. T and Graczyk, P. and Schneider, U. (2021). The Geometry of Model Recovery by Penalized and Thresholded Estimators, *HAL*.

Tibshirani, R. (2013). The lasso problem and uniqueness, *Electronic Journal of statistics*, 7, 1456–1490.